DOCUMENT RESUME

ED 336 095 IR 053 645

AUTHOR Trahan, Eric
TITLE Comparing Computerized and Paper Based Information Retrieval Systems in Library Public Services: A Meta-Analysis.
PUB DATE Dec 90
NOTE 40p.; M.L.S. Thesis, State University of New York at Albany.
PUB TYPE Dissertations/Theses - Masters Theses (042) -- Information Analyses (070) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Bibliographic Databases; *Card Catalogs; Comparative Analysis; *Indexes; *Information Retrieval; Information Technology; Library Automation; Meta Analysis; *Online Catalogs; *Online Searching; Relevance (Information Retrieval)

ABSTRACT

A meta-analytical literature review was performed on the literature in which computerized and paper based information retrieval systems were compared. Specifically, online catalogs were compared with card catalogs, and online bibliographic retrieval was compared with searching printed indexes. Studies which included information on relevance, precision, time, or costs of searching were selected. A total of 25 studies published between 1967 and 1989 met the selection criteria, producing a total mean effect size of -0.383. The analysis revealed that there were no significant differences between the two systems for the variables of relevance, time, or costs. The paper based systems were significantly superior on the precision variable. The variance in individual study results could not be explained by any of the factors that were included in the analysis. These factors included the publication date, publication mode, method of effect size computation, library environment, and search complexity. It is hypothesized that the variability in study methodology might explain the variability in study results. Specific recommendations are made for more standardized methods in future research in which information retrieval systems are compared. The individual study results are appended. (54 references)
(Author/MAB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED [illegible]

COMPARING COMPUTERIZED AND PAPER BASED INFORMATION RETRIEVAL SYSTEMS IN LIBRARY PUBLIC SERVICES: A META-ANALYSIS

by

Eric Trahan

Submitted in partial fulfillment
of the requirements for
the M.L.S. degree
at
The School of Information Science and Policy
The University at Albany
State University of New York

ISP 680: Seminar in Information Science & Policy
Professor Norman Hoyle
December 1990

12053645

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Eric Trahan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Abstract

A meta-analytical literature review was performed on the literature in which computerized and paper based information retrieval systems were compared. Studies which included information on relevance, precision, time or costs of searching were selected. A total of 25 studies published between 1967 and 1989 met the selection criteria. A total mean effect size of -0.383 was produced. The analysis revealed that there were no significant differences between the two systems for the variables of relevance, time or costs. The paper based systems were significantly superior on the precision variable. The variance in individual study results could not be explained by any of the factors that were included in the analysis. These factors included the publication date, publication mode, method of effect size computation, library environment and search complexity. It is hypothesized that the variability in study methodology may explain the variability in study results. Specific recommendations are made for more standardized methods in future research in which information retrieval systems are compared.

The computerization of library services, specifically of information retrieval, has been the most discussed and debated topic within library circles in recent years. Since the inception of computerized indexes in the early 1960s, countless books and journal articles have been published which discuss the advantages and disadvantages of automated information retrieval. Indeed, many new journals have been introduced, which have as their expressed purpose the reporting of information on this particular aspect of automation.

Though some authors state that librarians universally consider this automation an improvement (Lipow, 1989; Miller, 1989), others are just as vehement in opposing the new systems (Kusack, 1988a, 1988b). Those who favor the automation of information retrieval point to the many advantages which computerized systems should be able to offer. Librarians in this camp claim that automation will lead to improved productivity and error control; and also to increased speed, range and depth of service. More specifically, online systems are expected to increase the depth of indexing, provide multiple access points, allow for better updating and allow librarians to use materials that are not physically present. Those who favor traditional paper based systems point to the portability, the browsing capabilities, and the freedom from equipment that these systems offer (Lancaster, 1977, 1982; Lipow 1989). Unfortunately, however, most librarians are not basing their opinions on any real advantage offered by either system. There is evidence that

the positions taken by most librarians on this topic are based more on irrational opinions than on hard evidence (Kusack, 1988a, 1988b).

Whatever the opinions expressed by librarians, many libraries have already automated their information retrieval systems. Seventy four percent of academic libraries now have online catalogs (Epple & Ginder 1987), and online database searching has become commonplace (Medow 1988). These innovations have a direct impact on library users. Library researchers have been quick to administer surveys that attempt to assess users' opinions of these new technologies. These surveys seem to show that users overwhelmingly support the automation of information retrieval (Kranich, Spellman & Hecht, 1984; Lipow, 1989; Moore, 1984). Most of these researchers report that approximately 75% of those surveyed prefer the automated systems (California University, 1983; Ferguson, 1982; Lawrence, 1982; Markey, 1983), although results as high as 94% (Dowlin, 1980) and as low as 68% (Shuman, 1983), 64% (Pease & Gouke, 1982) and 16% (Edmonds, Moore & Balcome, 1989) have also been reported. Although these results appear to be conclusive, the methodologies used are not beyond reproach. Most are highly vulnerable to the hawthorne effect. Also, library patrons' expressed opinions may not be an accurate measure of their true satisfaction. Kranich Spellman and Hecht (1984) provide some insight into this problem. The results of their study showed that 63% of the subjects using the card catalog found the material that they were seeking, while only 35%

of the online group subjects successfully completed their searches. Yet, these researchers also reported that 75% of the subjects preferred the online catalog. These results are obviously inconsistent with one another.

Although the purposes of the authors of all of these research papers and opinion articles have been to clarify the issues involved in the computerization of information retrieval in libraries, the real result has often been a clouding of the issues. So much has been written on this topic, and so much conflicting evidence has been reported, that the sum effect is confusion rather than clarity. A further problem for researchers in this area has been the economic realities of library operations. If automated information retrieval systems were implemented as additions to library services there could be no argument that they provide for increased capabilities over traditional systems alone. However, for economic reasons, automated systems often replace rather than augment existing systems (Epple & Ginder, 1987; Lancaster, 1982).

There is, therefore, a need for studies which experimentally compare the merits of these two systems; and also for reviews that provide digests of all that has been written on the automation of information retrieval. This research project is designed to provide librarians with answers to these two needs. These goals will be accomplished through a statistical review of the experimental evidence that has been reported in the library literature which directly compares automated and traditional

information retrieval systems in libraries. This will be accomplished through a methodology known as meta-analysis.

Methods

Description of Meta-analysis

The methods used in meta-analytic research developed slowly during the 1950s and 1960s, as researchers sought to cope with the vast quantity of experimental data that was available for many research questions. These methods were first described as a distinct research methodology, and were first called meta-analysis, by Gene Glass in 1976. Other researchers, most notably Robert Rosenthal, have proposed alternate methods for integrating study findings (Bangert-Drowns, 1984; Glass, McGraw & Smith, 1981; Rosenthal, 1984). This project will use the principles described by Glass, with some modifications, as advocated by Bangert-Drowns.

Glass considers meta-analysis to be the incorporation of scientific and statistical methods into the practice of reviewing the literature. Meta-analytic literature reviews should be held to the same standards as primary research. This means that methodologies should be clearly described, results should be statistically analyzed and results should be replicable. The advantages of this type of research are many. The results are not dependent on the bias of the reviewer, and often a robust overall result can be obtained by combining the results of many inconclusive studies. Also, meta-analysis provides an

opportunity for evaluating the methods used in the primary research that is analyzed. Disadvantages, or criticisms, of meta-analysis include: that it lumps together data from studies done in different environments, that data from low quality studies is used, and that only published data is available for integration. Glass, McGraw and Smith (1981) provide rebuttals to each of these criticisms, and the methods used in this study will be designed to reduce the impact of these problems.

Data Collection

Studies which compared online information retrieval with paper based retrieval were located by searching **ERIC**, **LISA**, **ISA**, **NTIS**, **Library Literature**, **Dissertation Abstracts**, and the Online Information Retrieval Annual Bibliography, which is published in **Online Review**. To reduce the variability in the research environments, the study pool was limited to those works which compared online catalogs with card catalogs, or online bibliographic retrieval with searching printed indexes. Each study was required to report information on at least one of the following dependent variables: recall of relevant material, precision of the recalled set, the time necessary to identify each relevant hit, and the costs involved in identifying each relevant hit. It was necessary to reject many studies for reporting insufficient data. In order to be included, studies had to report one of the following forms of numerical results: means and standard deviations; recall or precision ratios, with the total number of relevants also reported; *F*, *t*, or chi square

statistics; or data from which any of these could be computed or estimated. Each study was also analyzed for its geographic location, year and mode of publication, library type, and search complexity. For the purposes of this analysis searching was defined as simple or complex depending upon whether boolean logic was used.

Data Analysis

For each study an effect size (es) was computed. Glass' definition of effect size is given in equation 1.

$$es = \frac{Me - Mc}{SD} \qquad (1)$$

Me in this equation is the mean of the experimental group, in this case the online group, Mc is the mean of the control or manual group, and SD is the standard deviation of the control group. Thus studies which show online searching to be superior will have positive effect sizes; and studies with results that favor manual systems will have negative effect sizes. The effect sizes for studies which reported means and standard deviations were computed according to this equation. It should be noted, however, that on the variables of time and cost it was necessary to change the sign of the result in order to keep the convention of having positive effect sizes for studies that favor automation. This was necessary since smaller means are superior for these variables. The effect sizes of studies that supplied

other statistics were computed using other formulas provided by Glass, McGraw and Smith (1981). Unfortunately, the effect sizes for studies which reported chi squares, recall ratios or precision ratios, could only be estimated, since these are non parametric statistics. Glass also provides formulas to perform these estimates.

An effect size was computed for each dependent variable of each study. These dependent variables include recall, precision, time per relevant citation and cost per relevant citation. An overall effect size was then computed for each study. Mean effect sizes (ES) were computed for each of the dependent variables, and for the overall study effect sizes. These ES were tested for statistical significance with t tests. In addition, the results were analyzed on the basis of the publication date, publication mode, library type, search complexity, and means of computing es. These analyses were made to explain the variability in individual study findings.

Results

Twenty five studies were located which met the selection criteria. The publication dates ranged from 1967 to 1989, with a mean of 1980. Eighteen of these studies were published in journals, three as ERIC documents, two as parts of conference proceedings, one as a research report, and one as a dissertation. A few of the studies were national in scope; nine states and four foreign countries were also represented. The foreign countries included the United Kingdom, Sweden, the Netherlands and Japan.

All sections of the United States were represented, including the northeast, southern, midwest, mountain and west coast states. For thirteen of the research projects the environment was an academic library. Eleven took place in special libraries and one in a public library. For ten of the studies the effect size was computed, for the remaining fifteen it was estimated. The search complexity could be defined as complex in thirteen studies and as simple in six. One study reported data for both simple and complex searches. For the remaining studies the search complexity could not be determined. All of the studies reported data on the recall of relevant material, eight on precision, eleven on time and seven on costs. The overall effect size most favorable to online searching was +4.750; that most favorable to traditional systems was -2.831. Negative effect sizes were obtained from nineteen of the papers; positive effect sizes from six. The data obtained from each study are reported in the appendix.

The mean effect size for each of these variables was computed. These results were: overall ES -0.383 with a standard deviation of 1.35; recall ES -0.503, standard deviation 1.47; precision ES -1.197, standard deviation 0.83; time ES +0.815, standard deviation 1.35; and cost ES -0.171, with 1.59 as the standard deviation. Statistical significance was obtained only on the precision variable. These results are summarized in table 1.

Table 1

Statistical Analysis of Effect Sizes

Variable	N	SD	ES	Statistic	Significance
Total	25	1.35	-0.383	t=1.41	no
Recall	25	1.47	-0.503	t=1.71	no
Precision	8	0.83	-1.197	t=4.09	p>.01
Time/hit	11	1.35	+0.815	t=1.99	no
Cost/hit	7	1.59	-0.171	t=0.28	no

Further analysis was done on the total mean effect size. The purpose of this analysis was to explain the variability in the individual studies' effect sizes. This analysis did not produce any statistically significant results. These results are summarized in table 2.

Table 2

Analysis of Variance Among Studies

Variable	N	SD	ES	Statistic	Significance
Year of Publication					
overall	25	1.35	-0.383	r=0.05	no
Computation of es					
computed	10	0.69	-0.305	t=0.22	no
estimated	15	1.65	-0.433		
Publication type					
journal	18	1.47	-0.259	t=0.71	no
other	7	0.93	-0.701		
Library type					
academic	13	0.76	-0.462	t=0.73	no
special	11	1.68	-0.066		
Search Complexity					
complex	14	1.53	+0.051	t=1.92	no
simple	7	0.88	-1.211		

Discussion

Analysis of Average Effect Sizes

Although most of these results do not show statistical significance, there is a strong practical significance. These results clearly indicate that computerized information retrieval systems do not, when taken collectively, offer any advantage over the traditional paper based systems. On the other hand, paper

based systems also have no clear overall advantage. These results have three important implications for libraries. The first is that libraries which are contemplating the automation of their public service information retrieval systems should not cite the improvement of information retrieval as a reason for pursuing automation. The automated systems analyzed in this review were superior to the manual systems only on the variable of time, and this superiority was not statistically significant. This is not meant to suggest that libraries should abandon computers. Automation has many advantages. There is little doubt that library technical services have become more efficient, and that cooperation among libraries has been enhanced. The overwhelmingly favorable reaction of library users to automation, as mentioned in the introduction, is also undeniable. Librarians should be aware, however, that the improvement of information retrieval in library public services is not one of the advantages offered by automation.

A corollary of this finding is that libraries which are striving to improve their public services should not see either automation or de-automation as the answer. Many of these studies reported that there was little overlap in the citations produced by the two systems. This indicates that both systems together may be the best alternative. This conclusion is supported by the research of Maciuszko (1987) and Caren and Somerville (1986). Maciuszko, who was not impressed with the retrieval of either system, states that "to abandon totally one system in favor of

the other may prove a disservice to the researcher" (Maciuszko, 1987 p309).

The second main implication of these results is that online systems have not met their potential. This conclusion has been reached by many other authors (Kusack, 1988b; Lipow, 1989). Again, the implication is not that we should abandon automated information retrieval, but that many enhancements will be necessary before its promise is fulfilled.

The final implication concerns the cost variable. Online searching is often claimed to be exceedingly expensive. Breen (1987) states that free online searching is economically unfeasible for most academic libraries; and reports that 73% of these libraries charge for online searches. This meta-analysis, however, does not support the assertion that online searching is significantly more expensive than paper based searching. East (1980), in an earlier review, reached the same conclusion. East also, however, provides a reason for the lack of significance; most studies are actually only crude comparisons, and don't take all aspects of the costs into account. This author has reached a similar conclusion. This conclusion was supported by Cohen and Young (1986). These researchers performed the only complete analysis of costs that was analyzed for this review. They found that print was cheaper for all databases analyzed, at both one year and five years of use.

Analysis of Variance Among Studies

The lack of significance in the analysis of the variance

among the individual study findings presents more of a problem. Analysis was made on the basis of year because of the often repeated assumptions that online systems are improving, and that, as computers become more commonplace, users will become more comfortable with their presence in the library. The relationship between publication year and study result, however, was virtually zero. Further evidence that users are not getting better at online searching is provided by Edmonds Moore and Balcome (1989). The subjects in this recent study were 10 to 15 year-olds. They were members of the so called computer generation. Yet, the effect size of this study was the most negative of all of those that were analyzed. This was also the only study among those analyzed in which the subjects preferred the card catalog. These results clearly indicate that user success with online systems has improved little since the inception of these systems 25 years ago.

The data was analyzed on the variables of publication type, and method of determining effect sizes, to check for a publication bias; and to insure that the estimations of effect sizes were accurate. Analysis on these variables is suggested by Glass McGraw and Smith (1981). Statistical insignificance on these variables is desireable, since significance would indicate that there was a problem with this research. The results on these variables, found in table 2, show that there is no apparent bias in the published literature on this subject. There was also no problem with estimating the effect sizes for studies that

reported insufficient data for computation. The average effect sizes for results that were computed and for results that were estimated were virtually identical.

Statistical significance was expected on the variable of search complexity. Research results reported by Havener (1988), and by Smille, Nugent, Sander and Johnson (1988), indicated that there was a difference in retrieval performance between simple and complex searches. Both of these papers reported data for both simple and complex searches; and both found that online systems were more advantageous for complex searches. Significant results, however, were not achieved on this variable in this meta-analysis, though this variable did produce a difference greater than that of any other variable. The average effect size for complex searches was +0.051; for simple searches it was -1.211. Significance may have been precluded on this variable because it was impossible to determine the degree of search complexity, and because complexity could not be determined at all for 5 of the studies.

The last variable on which variance was analyzed was library type. Analysis was performed on this variable because different types of libraries have widely different environments and widely divergent queries. The results from academic and special libraries were used in this analysis. The results from the one study done in a public library were not used. Significance was not achieved on this variable. This indicates that there was no difference between these two library types when comparing the

efficiency of search systems.

If the variance could not be explained by study date, effect size computation, publication type, or library type; and if it could only partially be explained by the search complexity, then how can we account for the considerable variance that did exist? There are at least two other possibilities. Unfortunately, analysis was not possible on these two variables. One possibility is the expertise of the searcher. It is known that end users and trained searchers produce very different results. Analysis was not possible on this variable because many of the researchers failed provide this information; and also because it was not uncommon for patrons to perform the manual search and librarians the online search.

The second alternative variable for explaining the variance in study findings is the quality of the methodology employed in the original study. The studies analyzed in this meta-analysis employed many different methodologies. Some of these methodologies were unbiased and well validated (Edmonds, Moore & Balcome 1989). Others were not validated at all, and blatantly favored one system or the other (Naber, 1985; Poynard & Conn, 1985). It should be noted that the effect sizes extracted from these and other questionable studies are among the most extreme of all of those that are reported in the appendix. It follows that much of the variance in individual study findings could possibly be accounted for by this variable. Analysis was not performed on this variable because it was not possible to

quantify the quality of each study's methodology.

Recommendations For Future Research

One of the advantages of meta-analysis mentioned in the description of this methodology was that it provides an opportunity for evaluating primary research methodologies. The importance of this evaluation is illustrated above. The variability in method quality may have been the factor that had the greatest effect on the variability in the study results. This section of this meta-analysis will contain recommendations for future studies which compare information retrieval systems. This will be done through an analysis of the methods used in some of the research reviewed here.

General Recommendations

One of the basic principles of all social science research is that robust, generalizable results can be obtained only from studies that employ multiple subjects. Comparative research which is performed with a single subject can provide information only on that one person. This should be obvious, and yet many of the studies analyzed for this review, including Miller (1968), Gill (1974), Santodonato (1976), Langley (1976) and Murphy (1985), were single subject studies. For this particular type of research it is also important to use multiple queries, since different questions may be more effectively answered in one system or the other. It would also be useful to perform separate analyses for questions of different complexity. There is evidence which suggests that the advantages of automation may be

better demonstrated by more complex questions (Havener, 1988; Smille, Nugent, Sander & Johnson, 1988). It is also important to randomly assign subjects to the experimental groups; or to have all of the subjects perform both searches. Most of the studies reviewed here which employed multiple subjects did adequately meet this requirement (Hartley, 1983; Havener, 1988).

Another very important requirement for research in general is that the methods must provide a valid answer to the research question. All of the studies analyzed here had the same basic research question: do paper based and computerized information retrieval systems differ in their retrieval effectiveness? Some of the methods used were effective in answering this question, but others were not. Naber (1985) and Poynard and Conn (1985) provide examples of methodologies that do not answer this question. Naber used an existing printed bibliography on water harvesting as his control or manual search. He then performed multiple exhaustive searches in many different online files, and compared the total number of online relevant results to the existing bibliography. This was published as proof that the online system was superior. All it really proved, however, was that an exhaustive search could produce a few more results than the existing bibliography; no existing bibliography could possibly contain all of the relevant citations. The methodology used by Poynard and Conn (1985), on the other hand, was slanted in favor of manual systems. These researchers performed a single MEDLARS search as their online search, and for their manual

result searched the contents pages of all of the journals that were known to publish papers on their subject. This methodology did not provide an answer to the research question. It actually showed that bibliographic searching could not substitute for an extensive personal knowledge of the literature.

The last general recommendation is that the two systems must be compared for similar levels of service. Some of these studies, for example Akeroyd and Rogers (1976) and Rogers (1985), used methodologies in which librarians performed the online searches, and students the manual searches. This is, of course, the situation that exists in many academic libraries. This methodology, however, can not be used to show that either system is superior, since the subjects in the two groups are not equivalent. Researchers that employ this methodology should not claim to be comparing retrieval systems, and their results can not be considered generalizable to other environments.

Recommendations for Determining Recall

The number of relevant citations recalled is obviously the single most important dependent variable when determining retrieval effectiveness. All of the studies analyzed in this meta-analysis compared the two retrieval systems on this variable. Most of these researchers used recall ratios as their descriptive statistic. Recall ratio is defined as the number of relevant hits retrieved over the total number of relevant citations in the database. This statistic is widely used in information retrieval research. It is not, however, the most

efficient or the most accurate unit of analysis. There are many serious problems with the computation of recall ratios. One is that, for most valid research methods, an extra and unnecessary computation is involved. All information retrieval research should be comparative; it is impossible to compute the absolute effectiveness of any system. All comparative research should use equivalent groups; that is, there should be equivalent numbers of relevant citations in the systems that are being compared. The denominator in the equation should therefore be the same for all of the systems in any well executed research. There is no reason to divide every result by the same total number. The computation of recall ratios, however, is not merely a waste of time. The denominator of the equation introduces an enormous confound into the analysis. Lancaster (1977) states that it is virtually impossible to adequately compute the number of relevant citations in a database. Researchers are therefore forced to estimate this figure. Different researchers, however, will produce different estimates; and those that are less assiduous will identify fewer total relevant citations for the database. In the formula for computing recall ratio we can see that this smaller denominator will result in a more impressive recall ratio. Thus we have produced the worst of all research situations; research that is poorly done will produce more significant results.

Recall ratios should therefore not be used in this type of research. Researchers should instead report the mean number of relevant citations produced by each system, along with standard

deviations. These statistics offer many advantages over recall ratios. They are easy to compute, and are universally recognized parametric statistics that allow for the computation of statistical significance. They also do not have the reliability problem that has been identified for recall ratios.

The number of relevant citations produced is important, however, only when comparing searches that are exhaustive. Many actual searches are not meant to produce all of the relevant material. Searches of this type should be evaluated not by the number of relevant hits produced but by the time expended in producing each relevant citation. This will be discussed further in the section on time.

Recommendations for Determining Precision

Precision ratios are a unit of little value. Lancaster (1977) identifies precision as a component of time; that is, the amount of time which the user needs to expend in determining the relevancy of the retrieved citations. This statistic is defined as the number of relevant citations produced over the total number of citations produced. Both of these figures are easily obtained. This statistic therefore does not have the serious problems that recall ratios have. However, it is still often poorly computed. Murphy (1985), for example, reported precision ratios that were seriously flawed. This researcher performed complex searches in various databases. She computed the precision ratios for the online searches in the usual manner; that is, the number of relevant citations produced by her search

statement divided by the total number of citations produced by the search statement. The precision ratio for the manual search was computed differently. The denominator of the equation was instead the total number of citations produced by the individual parts of the search statement. This resulted in a very inflated denominator, and therefore a very small precision ratio. Other researchers reported precision data that was flawed in the opposite direction (Elchesen, 1978). The subjects in this research project performed their own relevancy judgments as they performed the manual searches. All of the final results were therefore relevant. They therefore reported 100% precision for the manual searches. The online searches, however, were treated differently. Relevancy judgments were not performed during the online searches. All of the final citations were therefore not necessarily relevant. The precision ratios were then computed in the usual manner. This resulted in precision ratios that were biased in favor of the manual systems.

Since this confusion exists over the computation of precision ratios, it would probably be better not to compute them, but to include this component of the search results as part of the time variable, which will be discussed in the next section. However, if researchers feel that they have accurately determined precision ratios, and wish to include these ratios in their results, they should compute separate precision ratios for each subject and each question. They should then report the mean precision ratio for each system, along with standard deviations.

Statistical significance should also be reported. It is important to point out that researchers should not report precision ratios if they are also including this aspect of the search as part of the time variable. This would be including a single component twice in the data analysis.

Recommendations for Determining Time/Relevant Citation

As indicated in the previous sections, time is always an important variable, and is in some circumstances the most important variable. It is generally recognized that time should be measured as the time necessary to produce each relevant citation. Most of the research studies reviewed measured time in this way. This does not mean that there are no problems with measuring time. The most important consideration when determining time is that all aspects of the time used need be included for each search. This is because manual and automated systems differ greatly in time consumption; different aspects of each take different amounts of time. Thus the total time for each subject should include the preparation time, the search time, and where precision is not computed, the time needed to identify the relevant citations. This total time should then be divided by the number of relevant citations produced by the subject. Researchers should then report the means and standard deviations for each system.

Research into searches that are not exhaustive should have time per relevant citation as the primary variable. This can be accomplished by having searchers stop when they have reached a

pre-determined number of potentially relevant citations. The actual number of relevant results should then be determined. Time should then be computed as indicated above, and the means and standard deviations for the time per relevant cite should be reported.

Recommendations for Determining Costs/Relevant Citation

The cost of searches is also an important variable, and one that suffers from similar problems. East (1980), in an early review of cost comparisons, concluded that most studies did not directly compare the systems, but actually used crude estimations of the costs involved. Many researchers also did not include all aspects of costs in their analysis (Calkins, 1977; Elman, 1975; Huang & McHale, 1990). Lancaster (1977) identified 10 important aspects of cost analysis. These include start up costs, such as equipment and storage costs; and ongoing costs, such as materials and subscriptions. Staff salaries should be included only if this component was not included in the analysis of time. It is also important to report standard deviations. Many research studies, such as Cohen and Young (1986), were disqualified from inclusion in this meta-analysis because they did not report standard deviations.

Recommendations for Determining User Satisfaction

It was not possible to include data on satisfaction in this meta-analysis because few studies could be located which performed adequate analysis on this variable. Most of the studies mentioned in the introduction evaluated user satisfaction

with the two systems by asking a single question similar to: which do you prefer, the online catalog or the card catalog? User satisfaction with library services can not be measured by asking a single question shortly after a new and exciting service has been introduced. User satisfaction can be measured only with a multiple question assessment instrument that has been tested for reliability and validity. There is a pressing need in library research for such an instrument. Other guidelines for determining user satisfaction are provided by Lancaster (1977) and by Tessier, Crouch and Atherton (1977).

Recommendations for Future Meta-analyses

This is, as far as this researcher has been able to determine, the first time that meta-analytical procedures have been applied to library research. It is hoped that other researchers will see the value of this research procedure, and apply it to other research questions. There are surely other aspects of library research that could benefit from an unbiased statistical review; and from a critical analysis of the commonly used research methods. Future meta-analyses would also be useful on this particular research question, especially if analysis of the variance was possible on the variables of level of service and quality of methodology.

Conclusions

This meta-analysis has produced results that will be of value to librarians, to online system developers, and to library researchers. The main conclusion for librarians is that there is

little difference in retrieval effectiveness between paper based and computerized information retrieval systems. This is not to say that there are no differences between the two systems. Each offers unique advantages; and the wisest choice, if economically possible, would probably be to provide both services. The results of this research merely show that all statements which purport that either system provides generally enhanced information retrieval are based on something other than established fact.

Online system developers may be interested in the above conclusion. They should also note the lack of significance on the analysis which compared study publication date with effect size. Despite all claims to the contrary, there is no evidence that there has been any improvement in either the retrieval performance of online systems, or in the ability of searchers to effectively use these systems.

The conclusions for library researchers are two fold. The first is that meta-analysis can, and should, be used in reviews of the library research. The second conclusion is that the research into information retrieval systems has sometimes employed methodologies and statistical procedures that are flawed. Recommendations are provided to correct some of these problems. Some may consider these recommendations too rigorous. It should be noted, however, that robust and meaningful research results are not possible without rigorous research procedures.

Appendix

Individual Study Results

Study	Year	Pub. Type	Lib. Type	Search Type	Computation of es	Component es	Total es
Ohta	1967	J	A	na	computed	R +0.578	+0.578
Miller	1968	J	A	comp.	estimated	R -0.522	-1.276
						P -2.498	
Virgo	1970	J	A	na	estimated	R -0.340	-1.346
						P -2.351	
Bivans	1974	J	S	comp.	computed	R -0.047	-0.215
						P -0.047	
						T -0.513	
Gill	1974	E	A	comp.	estimated	R +0.019	-0.049
						P -1.640	
						T +0.967	
						C +0.457	
Goodliffe	1974	J	S	comp.	computed	R -0.851	-0.420
						T +0.009	
Michaels	1975	J	S	comp.	estimated	R -0.302	-0.302
Akeroyd & Rogers	1976	R	A	comp.	computed	R -1.601	-0.540
						P -0.978	
						T +0.949	
Langley	1976	J	A	comp.	computed	R -0.139	+0.618
						T +1.371	

Individual Study Results (continued)

Study	Year	Pub. Type	Lib. Type	Search Type	Computation of es	Component es	Total es
Santodorato	1976	J	S	simp.	computed	R -1.41	-1.410
Smith	1977	J	S	na	estimated	R -0.360	-0.185
						P -0.459	
						T +0.091	
						C -0.023	
Elchesen	1978	J	S	comp.	estimated	R +0.360	-0.004
						P -0.737	
						C +0.268	
Gouke & Pease	1982	J	A	simp.	estimated	R -0.971	-0.857
						C -0.750	
Hartley	1983	C	A	na	estimated	R +0.466	-0.198
						P -0.863	
Kranich	1984	E	A	simp.	estimated	R -1.035	-1.035
Murphy	1985	J	S	comp.	computed	R +0.076	+0.165
						T +0.487	
						C -0.071	
Naber	1985	J	S	comp.	estimated	R +4.750	+4.750
Poynard & Conn	1985	J	S	comp.	estimated	R -1.900	-1.900
Rogers	1985	J	A	na	estimated	R -3.358	+0.320
						T +1.981	
						C +2.341	

Individual Study Results (continued)

Study	Year	Pub. Type	Lib. Type	Search Type	Computation of es	Component es	Total es
Akaho, Banclai & Fujii	1986	J	S	comp.	estimated	R +0.149 T +4.352 C -3.419	+0.360
Frid	1987	C	A	simp.	computed	R +0.087 T -0.551	-0.231
Bernstein	1988	J	S	comp.	computed	R -1.570	-1.570
Havener	1988	D	A	both both simp. comp.	computed	R +0.135 T -0.182 -0.160 +1.090	-0.025
Reese	1988	J	A	simp.	estimated	R -1.950	-1.950
Edmonds, Moore & Balcom	1989	E	P	simp.	estimated	R -2.831	-2.831

Note.

Pub. Type (Publication Type): J = Journal, E = ERIC,
C = Conference, R = Report,
D = Dissertation.

Individual Study Results (continued)

Lib. Type (Library Type): A = Academic, S = Special, P = Public.

Computation of es: comp.= computed, simp.= simple, na = not available.

Component es: R = Recall, P = Precision, T = Time, C = Cost

References

Akaho, E., Bandai, A., & Fujii, M. (1986). Comparison of manual and online searches of Chemical Abstracts. Journal of Chemical Information and Computer Sciences, 26(2), 59-63.

Akeroyd, J., & Rogers, R. (1976). Online information retrieval: A comparison of manual and machine searching. (BLRR Report No. 5370)

Bangert-Drowns, R.L. (1984). Developments in meta-analysis: A Review of five methods. Ann Arbor, MI: Michigan University Center of Research on Learning and Teaching. (ERIC Document Reproduction Service No. ED 248 262)

Bernstein, F. (1988). The retrieval of randomized clinical trials in liver diseases from the medical literature: Manual versus MEDLARS searches. Controlled Clinical Trials, 9, 23-31.

Bivans, M.M. (1974). A comparison of manual and machine literature searches. Special Libraries, 65, 216-222.

Breen, M.L. (1987). Charging for online search services in academic libraries. College and Research Libraries News, 47, 400-402.

California University (1983). Results of a survey of users of the University of California prototype online union catalog. Part 1. Final Report. Berkley, CA: California University, Library Studies and Research Division. (ERIC Document Reproduction Service No. ED 231 400)

Calkins, M.L. (1977). Online services and operational costs. Special Libraries, 68, 13-17.

Caren, L., & Somerville, A. (1986). Online versus print sources in academic scientific and technical libraries: Supplement or replacement. Science and Technology Libraries, 6(4), 45-59.

Cohen, E., & Young, M. (1986). Cost comparison of abstracts and indexes on paper, CD-ROM and online. Optical Information Systems, 6, 485-490.

Dowlin, K. (1980). Users prefer online. Library Journal, 105, 1595-1596.

East, H. (1980). Comparative costs of manual and on-line bibliographic searching: A review of the literature. Journal of Information Science, 2, 101-109.

Edmonds, L., Moore, P., & Balcome, K.M. (1989). An investigation of the effectiveness of an online catalog in providing bibliographic access to children in a public library setting. (ERIC Document Reproduction Service No. ED 311 921)

Elchesen, D.R. (1978). Cost-effectiveness comparison of manual and online retrospective bibliographic searching. Journal of the American Society of Information Science, 29(2), 56-66.

Elman, S.A. (1975). Cost comparison of manual and online computerized literature searching. Special Libraries, 66, 12-18.

Epple, M., & Ginder, B. (1987). Online catalogs and shelflist files: A survey of ARL libraries. Information Technology and Libraries, 6, 288-296.

Ferguson, D. (1982). Public online catalogs and research libraries. Stanford, CA: Research Libraries Group. (ERIC Document Reproduction Service No. ED 229 014)

Frid, H.O. (1987). Perseverance as a factor for success at the online catalog. In. Online Information '87. 11th International Online Information Meeting: Proceedings (pp. 241-245). Medford, NJ: Learned Information.

Gill, E.D. (1974). A comparison of manual and computer searches of the chemical evolution and origin of life literature. San Jose, CA: San Jose State University. (ERIC Document Reproduction service No. ED 092 114)

Glass, G.V., McGraw, B., & Smith, M.L. (1981). Meta-Analysis in Social Research. London: Sage.

Goodliffe, E.C., & Hayler, S.J. (1974). Online information retrieval: Some comments on the use of Retrospec 1 in an industrial library. Aslib Proceedings, 26, 177-188.

Gouke, M.N., & Pease, S. (1982). Title searches in an online catalog and a card catalog. Journal of Academic Librarianship, 8, 137-143.

Hartley, D. (1983). A laboratory method for the comparison of retrieval effectiveness in manual and online searching. In Online Information '83. 7th International Online Information Meeting: Proceedings. (pp. 157-166). Medford, NJ: Learned Information.

Havener, W.M. (1988). The use of print versus online sources to answer ready reference questions in the social sciences (Doctoral Dissertation, University of North Carolina, 1988). Dissertation Abstracts International, 50, 565A.

Huang, S.T., & McHale, T.J. (1990). A cost effectiveness comparison between print and online versions of the same frequently used sources of business and financial information. In M.E. Williams (Ed.) National Online Meeting Proceedings 1990. Medford, NJ: Learned Information.

Kranich, N.C., Spellman, C.M., & Hecht, D. (1984). A study of user success with an online catalog. New York: New York University. (ERIC Document Reproduction Service No. ED 254 247)

Kusack, J.M. (1988a). Librarians and the information age: An affair on the rocks? Bulletin of the American Society for Information Science, 14(2), 26-27.

Kusack, J.M. (1988b). Librarians and the information age: Is reconciliation possible? Bulletin of the American Society for Information Science, 14(3), 27-30.

Lancaster, F.W. (1982). The evolving paperless society and its implications for libraries. International Forum for Information Documentation, 7(4), 3-10.

Lancaster, F.W. (1977). The Measurement and Evaluation of Library Services. Washington, DC: Information Resources Press.

Langley, P.R. (1976). A comparison between mail access computer and manual literature searching. RQ, 15, 229-232.

Lawrence, G.S. (1982). Users look at online catalogs: Results of a nationwide survey of users and non-users of online public access catalogs. Final report. Berkley, CA: California University, Library Studies and Research Division. (ERIC Document Reproduction Service No. ED 231 395)

Lipow, A.G. (1989). The online catalog: Exceeding our grasp. American Libraries, 20, 862-865.

Maciuszko, K. L. (1987). Hardcopy versus online searching: A study in retrieval effectiveness. In M.E. Williams & T.H. Hogan (Eds.), National Online Meeting: Proceedings - 1987 (pp.305-310). Medford, NJ: Learned Information.

Markey, K. (1982). Online catalog use: Results of surveys and focus group interviews in several libraries. Volume II of three volumes. Final Report. Dublin, OH: OCLC. (ERIC Document Reproduction Service No. ED 231 403)

Medow, C.T. (1988). Online database industry timeline. Database, 11, 23-27.

Michaels, C.J. (1975). Searching CA Condensates online vs. the CA keyword indexes. Journal of Chemical Information and Computer Sciences, 15, 172-173.

Miller, R.R. (1968). A study of searching the eye research literature. American Documentation, 19, 223-239.

Miller, W., & Gratch, B. (1989). Making connections: Computerized reference services and people. Library Trends, 37, 387-401.

Moore, C.W. (1981). User reactions to online catalogs: An exploratory study. College and Research Libraries, 42, 295-302.

Murphy, C.E. (1985). A comparison of manual and online searching of government document indexes. Government Information Quarterly, 2(2), 169-181.

Naber, G. (1985). Online versus manual literature retrieval. Database, 8(1), 20-24.

Ohta, M. (1967). A comparison of some demand subject searches: Machine vs. human. Bulletin of the Medical Library Association, 55, 408-415.

Pease, S., & Gouke, M.N. (1982). Patterns of use in an online catalog and a card catalog. College and Research Libraries, 42, 279-291.

Poynard, T., & Conn, H.O. (1985). The retrieval of randomized clinical trials in liver disease from the medical literature:A comparison of MEDLARS and manual methods. Controlled Clinical Trials, 6, 271-279.

Reese, C. (1988). Manual indexes versus computer-aided indexes: Comparing the Readers Guide to Periodic Literature to Info-Trac II. RQ, 27, 384-389.

Rogers, A.R. (1985). A comparison of manual and online searches in the preparation of philosophy pathfinders. Journal of Education for Library and Information Science, 26, 54-56.

Rosenthal, R. (1984). Meta-Analytic Procedures for Social Research. Beverly Hills: Sage.

Santodonato, J. (1976). A comparison of online and manual modes in searching chemical abstracts for specific compounds. Journal of Chemical Information and Computer Sciences, 16, 135-137.

Smith, R.G. (1977). Before you scrap the old ways compare retrieval systems: Manual vs. online. Online, 1(2), 26-27,51-59.

Shuman, B.A. (1983). Online and manual searches: A comparison. The Reference Librarian, 5/6, 173-180.

Smille, R.J., Nugent, W.A., Sander, S.I., & Johnson, D.M. (1988). A comparative assessment of paper based and computer based maintenance information delivery. San Diego: Navy Personnel Research and Development Center. (NTIS No. ADA 197 908)

Tessier, J.A., Crouch, W.W., & Atherton, P. (1977). New measures of user satisfaction with computer based literature searches. Special Libraries, 68, 383-389.

Virgo, J.A. (1970). An evaluation of Index Medicus and MEDLARS in the field of ophthalmology. Journal of the American Society of Information Science, 21, 254.